

WHAT IS CLAIMED IS:

1. A method for pre-processing entries in a directory listings, comprising:
 - receiving a first directory listings including one or more fields, the one or more fields populated with entries including one or more symbol strings;
 - receiving a second directory listings including one or more fields, the one or more fields of the second directory listings populated with entries including one or more symbol strings;
 - correlating entries in the one or more fields of the first directory listings with entries in the corresponding one or more fields of the second directory listings;
 - identifying entries, in the one or more fields of the first directory listings, which do not correlate with entries in the corresponding one or more fields of the second directory listings;
 - processing the identified entries using a rule set corresponding to the field in which the entry is located;
 - based on the rule set, determining a corresponding confidence level for the processed entries;
 - automatically modifying the processed entries having the corresponding confidence level meeting or exceeding a threshold; and
 - outputting the automatically modified entries for processing.
2. The method of claim 0, further comprising:
 - marking the processed entries having the corresponding confidence level below the threshold for operator confirmation.
3. The method of claim 2, further comprising:
 - presenting at least one of the marked entries to an operator using a graphical user interface;
 - presenting one or more rules from the rules set, corresponding to the field in which the at least one of the marked entries is located, to the operator using the graphical user interface;

receiving an operator's input selecting at least one of the one or more rules; and

processing the at least one of the marked entries in accordance with the operator's selection.

4. The method of claim 3, further comprising:

outputting the at least one of the marked entries processed in accordance with the operator's selection to an automated attendant.

5. The method of claim 3, further comprising:

outputting the at least one of the marked entries processed in accordance with operator's selection to a pre-processed listings database.

6. The method of claim 2, further comprising:

presenting at least one of the marked entries to an operator using a graphical user interface;

receiving an operator's inputs to manually modify the at least one of the marked entries; and

modifying the at least one of the marked entries in accordance with the manual inputs from the operator.

7. The method of claim 2, further comprising:

presenting one or more rules from the rule set, corresponding to the field in which the at least one of the marked entries is located, to the operator using the graphical user interface;

receiving an operator's input modifying the at least one of the one or more rules; and

processing the at least one of the marked entries in accordance with the modified rule.

8. The method of claim 0, wherein the processing step comprises:

selecting at least one of the identified entries;

based on the correlation with corresponding entries in the second database, determining whether the selected entry from the first database includes inverted symbol strings; and

if the selected entry is determined to include the inverted symbol strings, correcting the inversion in the selected entry.

9. The method of claim 0, wherein the processing step comprises:

selecting at least one of the identified entries;

based on the correlation with corresponding entries in the second database, determining whether the selected entry from the first database includes an abbreviation; and

if the selected entry is determined to include the abbreviation, expanding the abbreviation based on a closest correlation for the selected entry found in the second database.

10. The method of claim 0, wherein the processing step comprises:

selecting at least one of the identified entries;

based on the correlation with corresponding entries in the second database, determining whether the selected entry from the first database includes extraneous information; and

if the selected entry is determined to include extraneous information, removing the extraneous information based on a correlation for the selected entry found in the second database.

11. The method of claim 0, wherein the second database is an official postal office database.

12. Apparatus for pre-processing entries in a directory listings database comprising:

a reference database configured to store one or more fields, the one or more fields populated with entries including one or more symbol strings;

a rules database configured to store one or more rule sets; and

a processor configured to:

correlate entries contained in the directory listings database with entries in the corresponding one or more fields of the reference database,

identify entries in the directory listings database which do not correlate with corresponding entries in the reference database,

process the identified entries using the one or more rule sets from the rules database,

based on the one or more rule sets, calculate a corresponding confidence level for the processed entries, and

automatically modify the processed entries having the corresponding confidence level meeting or exceeding a threshold.

13. The apparatus of claim 12, wherein the processor to further output the automatically modified entries for processing.

14. The apparatus of claim 12, wherein the processor is configured with a word order normalizer that corrects word order of entries contained in the directory listings database.

15. The apparatus of claim 12, wherein the processor is configured with a street name expander that expands abbreviations of entries contained in the directory listings database.

16. The apparatus of claim 12, wherein the processor is configured with a township corrector that removes extraneous information from entries contained in the directory listings database.

17. The apparatus of claim 12, further comprising:

a confirmed listings database configured to store the automatically modified entries having the corresponding confidence level meeting or exceeding the threshold.

18. The apparatus of claim 12, further comprising:

a non-confirmed listings database configured to store entries that have the corresponding confidence level below the threshold.

19. A machine-readable medium having stored thereon a plurality of executable instructions, the plurality of instructions comprising instructions to:

- receive a first directory listings including one or more fields, the one or more fields populated with entries including one or more symbol strings;

- receive a second directory listings including one or more fields, the one or more fields of the second directory listings populated with entries including one or symbol strings;

- correlate entries in the one or more fields of the first directory listings with entries in the corresponding one or more fields of the second directory listings;

- identify entries, in the one or more fields of the first directory listings, which do not correlate with entries in the corresponding one or more fields of the second directory listings;

- process the identified entries using a rule set corresponding to the field in which the entry is located;

- based on the rule set, determine a corresponding confidence level for the processed entries;

- automatically modify the processed entries having the corresponding confidence level meeting or exceeding a threshold; and

- output the automatically modified entries for processing.

20. The machine-readable medium of claim 19 having stored thereon additional executable instructions, the additional instructions comprising instructions to:

- mark the processed entries having the corresponding confidence level below the threshold for operator confirmation.